View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by UUM Repository

International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 4 (2012) pp. 092-100
© MIR Labs, www.mirlabs.net/ijcisim/index.html

Usability Evaluation of SatNav Application on Mobile Phone Using mGQM

Azham Hussain[1] and Maria Kutar[2]

[1] Salford Business School, University of Salford,
Greater Manchester M5 4WT, United Kingdom
A.Hussain6@edu.salford.ac.uk

[2] Salford Business School, University of Salford,
Greater Manchester M5 4WT, United Kingdom
M.Kutar@salford.ac.uk

***Abstract*: SatNav systems are becoming increasingly popular, enabling drivers to locate and drive to their desired destination. With high demand for the features SatNav provides, these applications are now also available in mobile phones. This gives rise to the question of whether the app is usable when installed in a mobile phone. This paper will examine the usability of SatNav apps inside the mobile phone. We employ a mobile Goal Question Metric model (mGQM) to evaluate the usability of such applications by implementing usability test for objective measures, and questionnaire and interviews to assess subjective measures. Results indicate that most usability problems on SatNav apps in mobile are influenced by unique features of the mobile phone.**

Keywords: Mobile, Usability, Evaluation, SatNav System and GQM.

I. Introduction

Usability is commonly understood as a broad notion indicating the quality-in-use of interactive systems [1] and [2]. There are many measures of usability, and these include task completion time, error rates, time taken to key in the data, as well as subjective satisfaction, for example enjoyment, ease of use, safety and so forth. The word "usability" also refers to methods for improving ease-of-use during the early design process [3] and [4]. Focusing on usability and user experience is a key element in creating successful high-quality applications. However, the novelty of mobile applications (apps) and the unique features of mobile devices become key challenges in usability measurement activity.

SatNav apps (such as TomTom, Garmin and CoPilot Live) are becoming increasingly important to all kind of drivers, especially those who deliver goods, and are becoming compulsory to fire service and ambulance personnel. These apps can now be installed inside a mobile phone if the phone has a Global Positioning System (GPS) receiver. SatNav apps inside the mobile phone are different from the stand alone SatNav device. By using SatNav apps inside mobile phone, we are still able to make and answer calls, and simultaneously tilting some mobile phones enables us to change the screen orientation to portrait or landscape. Furthermore, the screen size of mobile phones is generally smaller than for standalone devices, and these differences give rise to the question of whether these apps are usable in the mobile phone.

The unique features of mobile devices and wireless networks pose a number of significant challenges for examining usability of applications on mobile phones. They are mobile context, multimodality, connectivity, small screen size, different display resolutions, limited processing capability and power, and restrictive data entry methods [5] and [6]. The features will influence usability of the app on a mobile phone, for example slow processing time or response time. In addition, users might not be satisfied with the application if the display resolution is low and the screen is too small.

Several studies provide methods to measure usability; however some limitations occur when they are applied to mobile phones [7]. Thus, we have developed a Mobile Goal Question Metric (mGQM) model to evaluate mobile apps. The model contains usability metrics to assess quantitative and qualitative measures of mobile phone apps. We developed the model based on the original GQM approach by Basili [8]. A standard GQM model is a hierarchy structure starting with goals, in which these goals are refined into questions in the later stage before metrics for each question are developed. We provide the objective and subjective metrics in section 3.

The objective of this paper is examine usability problems of SatNav apps on mobile phones. We used the mGQM model to evaluate them. We employ two SatNav apps that are installed in mobile phones as a test case to find out usability problems for both apps. In the following section (section 2), we examine previous studies on mobile phone evaluation and the background of SatNav apps in mobile phone. We provide details of the mGQM model in section 3, and Section 4 describes the method used to conduct the study, followed by the presentation of results in Section 5. Finally, we discuss conclusions arising from the study and provide recommendations for further work in Section 6.

II. Evaluation of Mobile Phone Apps

Mobile apps can be referred to as software systems operating on mobile devices [9]. The novelty of mobile apps and the unique features of mobile phone devices are becoming key challenges in usability evaluation. Recent developments in technology, such as GPS receivers and compass embedded into mobile phones, create a wide range of research opportunities. Primary areas of research on mobile phone include designing usable interfaces [10]; how usability is measured [11], or the relationships between the technology, work tasks and the context of work [12]. Understanding the usability of mobile apps has been widely discussed, for example, navigation of complex information on small screens [13]; tactile feedback [14]; techniques for assessing mobile usability[11, 15, 16], texting [17], and how to conceptualize mobile usability [18].

Usability evaluation methods refer to the techniques employed to carry out usability evaluation, such as usability testing, focus groups and interviews [19]. All these methods have been used by many researchers to evaluate usability and each method has its own distinct advantages and disadvantages depending on the objective of the study. Different evaluation methods have emerged and contributed to the evolution of usability evaluation, giving software development organizations a wide collection of techniques that fit specific development projects [20]. Two major methodologies to evaluate mobile applications are laboratory experiments and field studies [21]. In a laboratory experiment, human participants are required to accomplish specific tasks using a mobile application in a controlled laboratory setting, while a field study allows users to use mobile applications in the real environment.

A number of models for measurement are available for reference; for instance, Quality in Use Integrated Measurement (QUIM) developed by Ahmed et al. [22]. QUIM is a consolidated model for usability measurement and metric; and also appropriate for users who have no or little knowledge of usability. The model consists of 10 factors which are subdivided into 26 criteria. For the measurement of the criteria, the model provides 127 metrics. The model is used to measure the actual use of working software and identifying the problem; however, the model is not optimal yet and needs to be validated.

A number of metrics based models and tools for evaluating usability have been proposed over last 15 years. One such model is Metrics for Usability Standards in Computing (MUSiC) developed by Bevan & MacLeod [2]. MUSiC is a project concerned with defining measures of software usability and was integrated into the original ISO 9241 standard. Examples of specific usability metrics in the MUSiC framework include user performance measures, such as task effectiveness, temporal efficiency, and length or proportion of productive period. However, a strictly performance-based view of usability cannot reflect other aspects of usability, such as user satisfaction or learnability. Software Usability Measurement Inventory (SUMI) develop by Kirakowski & Corbett [9] is a part of the MUSiC project. SUMI was developed to provide measures of global satisfaction of five more specific usability areas, including effectiveness, efficiency, helpfulness, control, and learnability. Another MUSiC project related to software tool development is the Diagnostic Recorder for Usability Measurement [10] developed by Macleod & Rengger [11]. This project concerns with the analysis of user-based evaluations and delivery of these data to the appropriate party, such as a usability engineer. The Log Processor component of DRUM is the tool concerned with metrics. It calculates several different performance-based usability metrics including 1) Task time 2) Snag, help, and search times 3) Effectiveness 4) Efficiency 5) Relative efficiency and 6) Productive period.

In addition, the Automated Interface Designer and Evaluator (AIDE) that was developed by Sears [12] focuses on evaluating static HTML pages according to a set of predetermined guidelines about Web page design. AIDE is a software tool that is able to generate alternative interface layouts and evaluate some aspects of a design. Among things that are concerned in these guidelines are the placement and alignment of screen elements for example text, buttons, or links. There are two metrics to be evaluated in the design which are a task-sensitive metric and a task-independent metric. Task-sensitive metrics incorporate task information into the development process which may ensure that user tasks guide the semantics of interface design. Task-independent metrics tend to be based on principles of graphic design and help to ensure that the interface is aesthetically pleasing. The AIDE tool can measure a total of five different usability metrics, including efficiency, alignment, horizontal balance, vertical balance, and designer-specified constraints.

Subsequently, another model that deals with the analysis of the quality of use for interactive devices were introduced which is The Skill Acquisition Network (SANe) by Macleod & Rengger [11]. This approach assumes a user interaction model that defines user tasks, the dynamics of the device, and procedures for executing user tasks. Specifically, a task model and a device model are simultaneously developed and subsequently linked. After that, user procedures are simulated within the linked task-device model. A total of 60 different metrics are described in this framework, of which 24 concerns with the quality measures. Scores from the latter are then combined to form a total of five composite quality measures including: Efficiency, Learning, Adaptiveness, Cognitive workload, Complexity and Effort for error correction.

It can be seen that metric based approaches have been frequently taken to evaluate usability, indicating that this is a fruitful direction for research. However, all of the models outlined above have been developed for and are appropriate for desktop application systems, but none are aimed at mobile devices or mobile apps and so do not deal with the specific characteristics of interaction with mobile devices.

A study on the challenges and issues of mobile applications by Zhang [9] lists nine usability attributes and measuring variables as a part of their studies. All the generic attributes as listed in Table 1 were collected and compiled from existing usability studies such as Danesh et al., [23]; Frokjaer, Hertzum, & Hornbaek, [24]; Nielsen, [25] and Ziefle, [26]. In contrast, Gafni [27] introduced the usability quality characteristic for a mobile wireless information systems. The study focuses on the development of a new metric and all metrics were validated theoretically and empirically at least by one of four different experiments performed in diverse devices. However, the devices used in the experiment are quite aged and the model needs to be updated to provide new metric for new mobile phone. A study by Terrenghi [28] shows a detailed usability metric for mobile devices by refining the usability characteristic from ISO /IEC 9216-1. The study focuses on new issues and encounters usability requirements for mobile computing scenarios; however they are not yet fully validated.

Few studies can be found which focus on the evaluation of SatNav apps and this includes a study by Bernett [29] that lists three important improvements for the interface of SatNav systems, which are voice, display and control. Burnett also suggested improving timing on the final approach to manoeuvres and the ability to 'block out' those functions which are too demanding for drivers when the vehicle is in motion. Blandford, et al. [30] used a SatNav system as a case to produce a methodology to evaluate the usability of an interactive system in terms of the modalities of interaction. Both studies focus on SatNav systems as standalone systems used inside cars, and on the other hand, our study will focus on evaluation of SatNav system installed inside mobile phone.

Attribute	**Description**
Learnability	How quickly users can improve their performance levels.
Efficiency	How fast users can accomplish a task
Memorability	Level of ease with which users can recall.
Errors	Counting the number of mistakes that users make.
User satisfaction	Reflects the attitude of users toward using a mobile application.
Effectiveness	Defined as completeness and accuracy.
Simplicity	The degree of comfort with which users find a way to accomplish tasks.
Comprehensibility	Measures how easily users can understand content presented.
Learning performance	Measures the learning effectiveness

Table 1. Usability Attributes by Zhang [9]

III. Mobile Goal Question Metric (mGQM)

One approach to the measurement of software development is Goal Question Metric (GQM). The GQM approach was developed by Basili et al. [8], and it becomes a de facto standard for the definition of measurement frameworks [31]. The approach is successful for the reason that it is adaptable to many different organizations and environments, demonstrated by the large number of companies that have employed it (e.g. Philips, Siemens, NASA) [32]. Although the main interest for measurement activities in the original GQM is a software project, the underlying concepts are generic and applicable in many measurement setting and also provides a practical approach for bounding any measurement problem [33], [34] and [35].

The need for a generic evaluation model is increasing [36]. The flexibility of the GQM model suggests that is would be suited to this domain, and so we have followed the GQM method to develop generic usability metrics for mobile apps. The goal for the model is the quality characteristics of usability factors. We created the questions by carefully refining the goal into several questions. Finally, we develop a set of metrics that provide the information to answer those questions. The complete mGQM model can be viewed at [37].

Determining the goals is a very important factor before developing the model to ensure reliability and adaptability. Inaccurate goals will affect the quality of questions and metrics. Hence, we ensure effective goals are created by reviewing the literature related to evaluation and measurement. We searched five top journals in HCI area and one conference proceeding from years 2007 until 2010. The journals and conference proceeding selected are shown in Table 2. To select the goals from selected journals, we planned and conducted a systematic literature review following the Kitchenham's procedure [38]. All of the goals selected in this model are based on usability standards by [1]. As a result, six usability characteristics act as goals for mGQM model.

Journal & Proceeding / Year	2007	2008	2009	2010
TOCHI	0	4	4	1
HCI	1	1	0	1
IJHCI	3	1	1	5
IJHCS	3	2	6	4
IJMHCI	-	-	0	0
MobileHCI	0	0	1	5
Total	7	8	12	16

Table 2. Paper Selected from Journal & Proceedings

Based on the goals created, we the generated the questions to assess each goal. The questions constitute the basis for quantitative and qualitative metrics definition. Finally, a set of metric has been produced to provide information to answer the questions developed. Table 3 below summarizes the mGQM model and shows the usability metric to measure mobile apps.

Goals	**Questions**	**Metrics**
Simplicity Accuracy Time Features Safety Attractiveness	-Is it simple to key-in the data? -Is the output easy to use? -Is the application easy to learn? -Is the application accurate? -How many tasks are successful in a given time? -How much time taken to complete a given task? -How much time taken by user to learn? -Does the application provide appropriate help? -Does the application provide voice assistance? -Does the application provide automatic update? -How the users feel when using the application? -Is the application secure to use while driving? -Are users happy with interface? -Are users familiar with the user interface?	-Time taken to key-in the data -Satisfaction with virtual keypad -Satisfaction with help -Satisfaction with screen size optimization -Satisfaction with output -Time taken to install -Satisfaction with installation process -Time taken to learn -Number of mistake while learning -Number of error -Time taken to complete given task -Time taken to response -Time taken to connect to network -Satisfaction with menu button -Satisfaction with voice assistance -Number of voice assistance -Number of system resource display

	-No of request to update the apps -Enjoyment -Signal strength -Satisfaction with contents -Satisfaction with virtual joystick -Finding help -Satisfaction with interfaces -Safety while driving -Strain injury or stress. -Satisfaction on learning process

Table 3. Usability Metric for Mobile Apps

The metrics are separated into quantitative measures and qualitative measures to assess the usability of mobile apps. Further details of the model itself are available in [37]. The model is applied by performing usability tests to obtain quantitative data with the qualitative data obtained through questionnaires or interviews.

IV. Method

The purpose of this experiment is to review the usability of SatNav apps in terms of the way in which humans interact with the app. The usability data is collected and analyzed by following the mGQM rules for example objective data collected first follow by subjective measures or vice versa. Results from the experiment can provide an indication of whether SatNav apps are usable in mobile phones. The experiment is divided into two parts; first we collect the objective data through usability test, and then we collect subjective data via questionnaire and interview to assess the perception of participants on SatNav apps, as recommended by Nielsen [25]. We used the TomTom One and CoPilot Live SatNav system installed in an O2 Orbit mobile phone device and the experiments were conducted inside a car in order to mirror the way such apps are used in practice. However, participants did not drive the car during the study for safety reasons.

A survey by Kjeldskov [11] showed that 71% of all mobile HCI evaluations are done in the laboratory and only a few conventional usability test were customized to meet arising challenges of mobile application evaluation. This may be due to data collection techniques such as think aloud, video recording or observations being difficult in the field. Evaluation in the lab has many advantages such as controllable situation condition and reproducibility. However, the drawback of lab experiment is the lack of realism.

The similarities and differences between field and lab-based evaluations of mobile application are beginning to be explored. Several comparisons have been made to observed the different on interaction behaviours in the laboratory and in the field settings for instance by Baillie [39] and Pirhonen [40]. They compare between field and lab and conclude that it is worthwhile carrying out evaluations in the field, even though it is problematic due to difficulties in capturing screen content and the interaction between the user and the mobile device. This study will be mix due to the requirement of satellite signal by SatNav system, however the car will not moving for safety reasons.

The number of people to be included as participants in a study depends on time, money and effort to analyse the data [41]. In this study, twenty participants were recruited as suggested by Landauer [42] and Nielsen [25], and six of them were not familiar with the application to purposely obtain the data related to “learnability”. A mix of participants including novice and more experienced users provides different perspective and feedback while using the application [25]. Participants had a range of experience from novice to expert, and included men and women aged from 20 to 39. Participant’s profiles were described in Table 4. All participants were asked to complete five tasks, and they were given time to explore and learn the apps before continued to complete all tasks.

Task 1: Set the application to go to a specified address. This is the main objective of Satellite Navigation System and it is essential that this is represented in the study.
Task 2: Set the application to go to ‘Sheffield City Centre’. This task tests whether users can easily navigate using different sub menu. The task uses different a sub menu section compared to task one.
Task 3: Add a specified address into favourites.
This task utilises a different menu section within the main menu.
Task 4: Change the measurement unit from Miles (Imperial) to Kilometre (Metric). This task represents one of the main menus in SatNav system. This inclusion of this task ensures participants utilise all menus in SatNav.
Task 5: Delete a specified address from favourites: This tasks tests whether participants are able to locate a deep sub menu.

Participants	Age & Gender	Experience	Education
P1	33, Male	Experienced	Post Graduate
P2	30, Female	Novice	Post Graduate
P3	34, Male	Experienced	Post Graduate
P4	29, Male	Experienced	High School
P5	39, Male	Novice	Post Graduate
P6	35, Male	Experienced	Post Graduate
P7	31, Female	Experienced	Post Graduate
P8	35, Male	Experienced	Graduate
P9	28, Male	Novice	Post Graduate
P10	31, Male	Experienced	High School
P11	30, Male	Experienced	Graduate
P12	35, Male	Experienced	Post Graduate
P13	20, Male	Experienced	High School
P14	29, Female	Novice	Graduate
P15	34, Male	Experienced	Graduate
P16	31, Male	Novice	High School
P17	30, Male	Experienced	Graduate
P18	35, Male	Experienced	Post Graduate
P19	20, Male	Experienced	High School
P20	29, Female	Novice	Graduate

Table 4. Participant Profile

The data collected in this study were then compared with data collected from an expert user. The procedure for data collection for the expert user is the same as for the other participants. We selected the expert user based on experience using the apps for more than a year using the same device as in the experiment. Participants were grouped into two. The first group consisted of ten participants (P1 to P10) using the

TomTom app and the other ten used CoPilot apps. Both apps were installed in an O2 Orbit phone. Much of the data collected denotes periods of time since time taken reflects the effectiveness of the application as detailed in the metric. In addition to time taken, other data collected includes the number of errors, help assistance and the number of system resources displayed.

Subjective data was obtained using a 9-point Likert measurement scale; from one to nine, similar to Questionnaire for User Interface Satisfaction [13] originated by Chin et al. [43]. We used the rules contained within the mGQM model to create forty questions and the questions focused on user satisfaction of mobile phone apps. These include questions on satisfaction with aspects such as the virtual keypad, menu buttons, text size, voice assistance and signal strength. Finally, a post-study interview session was conducted to expand the understanding of user satisfaction while using SatNav apps. Participants were interviewed using a semi-structured interview technique to ensure participants are able to talk freely on SatNav apps in informal environment. This provides an opportunity to discuss areas that were not previously covered in the experiment and questionnaire.

V. Results and Discussions

Results are presented separately for objective and subjective data. We conducted usability test to record usability problem for both apps and also compare the result with expert user to check any significant differences between the two apps. Several pictures below were taken while the usability tests were being conducted.

A. Objective Measures

For objective measures, we collected the data and summarized it for each metric. Table 6 shows the mean, medium and mode value for each metric across the five tasks (we use the label from “A to “K” to represent the metrics). Data collected from expert user is also represented to illustrate the difference in values. The expert user was allowed to learn, and time taken to learn was recorded for the reason that the expert user may also need to refresh on how to use the SatNav app on mobile phone.

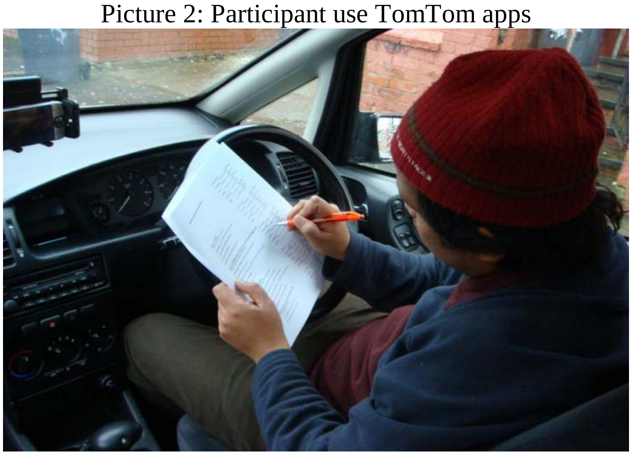
It is evidenced that the measures with greatest difference in time are A (time taken to key in the data) and C (time taken to learn). The time differences between the expert and other participants are the greatest because two novice users in this experiment needed extra time to learn and to become familiar with the app. They were not familiar with the virtual keypad and needed extra time to key-in the data.

We compare the performance between TomTom and CoPilot using similar data above to find out the more effective SatNav apps on mobile phone as shown in figure 1. The comparison method undertaking in this study is similar to [7] to determine whether one app is more usable than another. The comparison shows that all measures on CoPilot are greater than measures on TomTom except D (Number of mistake while learning), and this indicates that participants needed additional time using CoPilot to complete the task compared to TomTom. Figure 1 clearly indicates that TomTom inside the mobile phone is quicker compared to CoPilot. However, area under discussion is subject to measure for human interaction only.

Picture 1: Participant used SatNav Apps

Picture 2: Participant use TomTom apps

Picture 3: CoPilot Apps in O2 Orbit



Picture 4: Participant Answer questionnaire

Usability Metric	TomTom One				Co Pilot			
	Expert	Mean	Median	Mode	Expert	Mean	Median	Mode
A - Time taken to key-in the data	33	45.8	49.8	50	46	56.9	50	50
B - Time taken to install	181	181.6	182	182	188	189.7	190	187
C - Time taken to learn	12	44.2	47	47	22	60.5	61	61
D - Number of mistake while learning	0	1	1	1	0	0	0	0
E - Number of error	0	1	1	1	0	3	3	1
F - Time taken to complete given task	54.4	68	69	69	111	113.4	121	121
G - Time taken to response	9.6	10.36	10	10	17	17	18	18
H - Time taken to connect to network	30	35.4	34	34	55	53.8	55	48
I - Number of voice assistance	1	1	1	1	1	1	1	1
J - Number of system resource display	2	2	2	2	2	2	2	2
K - Number of request to update the app	0	0	0	0	0	0	0	0

Table 6. Result for Objective Measures

B. *Subjective Measures*

Results for subjective measures were obtained by analyzing the questionnaires and interviews sessions completed by participants.

Overall results show that users were satisfied with the application and they felt that the application was very helpful. However, most of the participants were undecided whether they were satisfied with the "Virtual Joystick" as the application was not using virtual joystick.

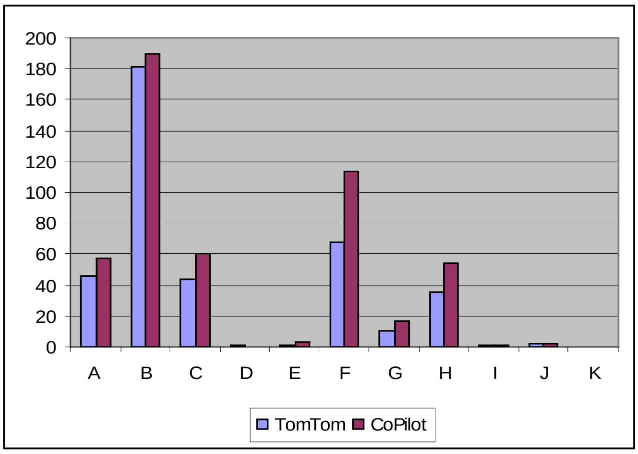


Figure 1. Comparison between TomTom and CoPilot

Table 7 summarizes the results given by participants. We matched the questions from the questionnaire with mGQM measures (for example questions 7 and 38 are similar to "virtual keypad" in mGQM) as shown in second column. The third column (score) is the mean score for the questions in second column. We used a 9-point likert scale where a higher score indicates greater satisfaction. Overall results show that the participants were satisfied with the application except for the virtual keypad. Interviews indicated that they were not satisfied with the keypad for the reason that the keypad is too small and they needed to use the stylus to key-in the data. Participants also commented on the letter arrangement on keypad. They suggested the keypad should follow standard arrangement similar to computer keyboard. We were unable to measure "Virtual joystick" because the tested apps do not use a virtual joystick.

Subjective Measures	Question No	Score
A: Virtual keypad	7,38	5.3
B: Help	13,15,16,27	7.1
C: Screen size optimization	9,31,33	6.8
D: Output	24,39	7.1
E: Installation process	28,29	7.4
F: Menu button	8,40,21	6.5
G: Voice assistance	17,25	7.8
H: Enjoyment	1,2,3,32	6.9
I: Signal strength	36,37	6.8
J: Contents	4,5,6	6.6
K: Virtual joystick	38	N/A
L: Finding help	14,22	7.1
M: Interface	10,11,12,30	6.9
N: Safety while driving	34	7
O: Strain injury or stress.	26,35	7.1
P: On learning process	18,19,20,23	7.1

Table 7. Summarized result for Questionnaire

We also present the data in diagrammatic form, shown in figure 2. The letters A to P represent the subjective measures in Table 7.

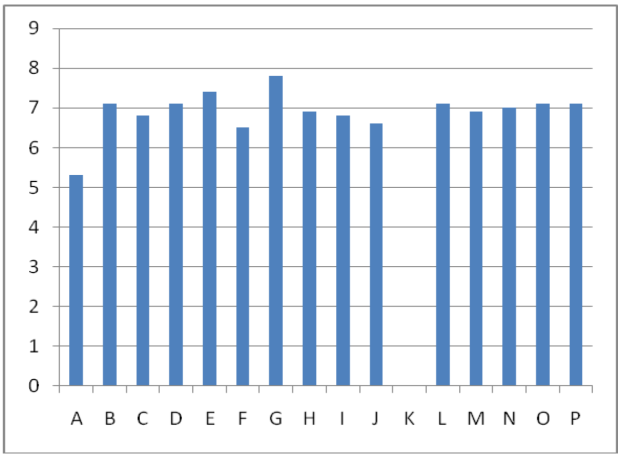


Figure 2. Result from Questionnaire

C. Interview Session

The third instrument used in the usability evaluation was an interview session with each participant in order to obtain opinions and perceptions while using the SatNav apps. The focal point of having an interview session is to get the subjective measures and to act as a support to the questionnaire, enabling participants to fully describe their opinions. The interview session was carried out after participants had completed the questionnaire.

The first question is concerned with the overall satisfaction on using the application. All participants agreed that the application is useful plus helpful, and in addition, that they would be keen to recommend the application to friends. This question reflects the satisfaction on the content of the application as well as enjoyment. The second question related to the arrangement of menu, submenu and the use of icons as menu items. This question was intended to determine if participants were satisfied with the interface of the application. All the participants were generally satisfied with the menu arrangement and picture buttons except one participant who suggested that the number of menus for applications being installed in mobile devices should be fewer compared to in standalone devices due to the small screen of the mobile phone.

The next question asked about possible improvement to the application. Five participants agreed that the application still has space for improvement and the other five had no suggestion. Two participants suggested that the application should have short cuts because they felt some of the task take too many steps to complete. Two participants suggested that the screen size should be extended to a more appropriate size, while one suggested reducing the connection time to satellite. Both suggestions are related to device issue and we do not discuss the device issues in this paper as they are beyond the scope of the evaluation.

In addition, all participants agreed that voice assistance is an important element in mobile applications; however one participant was not satisfied with the amount of voice assistance while using the application for both apps. This statement does not reflect any subjective measures as the amount of voice assistance was classed as an objective measure. However, we realize that the voice assistance appeared only one time for whole session. Since the car was not moving we were unable to obtain more voice assistance for example an instruction when the car approaching junction or roundabout.

The final question related to whether there is a need for such applications being installed into mobile phones. Only one participant did not agree and proposed to separate the application from the mobile phone itself, indicating that a mobile phone is specifically used to communicate with other people and no other applications should be installed on it. This situation did not reflect any subjective metric since the other nine participants agreed to have apps inside the mobile phone

VI. Conclusions

We have conducted a study to examine the usability of SatNav apps installed in mobile phones. Despite the fact that overall result shows that users were satisfied with the application and they indicated that the application was very helpful, several issues arose which need attention.

1) The virtual keypad should be enlarged and follow the standard “QWERTY” arrangement found on a computer keyboard. Pictures 5 and 6 show the standard computer keyboard and virtual keypad.

2) The apps should optimize the tiny screen because some of the screens still have space to put other menus and this would shorten the task.

3) Increase the amount of voice assistance as several participants mentioned that extra voice assistance would be helpful. This would be beyond the standard voice functions which provide direction.

Picture 5. Standard keyboard

Picture 6. Small virtual keypad

Conducting the study has revealed some limitations in the design of the study. Prior to conducting the test, we thought that it would be useful to measure the automatic update as part of the metric; unfortunately the update alert did not appear at all while conducting the usability test for all participants. Other issues include time taken to connect to satellite; connection time will be longer if the app is running for the first time. Furthermore, changing the location will also mean that it takes a longer time to connect to satellite.

The relatively close correspondence between the results for objective and subjective measures provides an indication that the metric itself is appropriate for use in evaluating the usability of mobile apps, although additional work is required to explore this in more detail.

For future tests, we will consider the participants to drive the car while conducting the test. In addition, there is a need to assess the extent to which results may have been influenced by the particular hardware and operating system used in this initial study. Further work will be considered using different operating systems and different devices to examine whether this leads to different results.

References

[1] ISO, "INTERNATIONAL STANDARD: ISO 9241-11(Guidance on usability)," Geneva 1998.

[2] N. Bevan and M. MacLeod, "Usability measurement in context," *Behaviour & Information Technology,* vol. 13, pp. 132 - 145, 1994.

[3] T. Yonglei, "Work in progress - introducing usability concepts in early phases of software development," in *Frontiers in Education, 2005. FIE '05. Proceedings 35th Annual Conference*, 2005, pp. T4C-7-8.

[4] J. Miller, "Usability Testing: A Journey, Not a Destination," *Internet Computing, IEEE,* vol. 10, pp. 80-83, 2006.

[5] I. Kondratova and I. Goldfarb, "M-learning: Overcoming the Usability Challenges of Mobile Devices," in *Networking, International Conference on Systems and International Conference on Mobile Communications and Learning Technologies, 2006. ICN/ICONS/MCL 2006. International Conference on*, 2006, p. 223.

[6] Nokia, "Getting Started with Mobile Design," Nokia Corporation Jun 5, 2008 2008.

[7] A. Echtibi, M. J. Zemerly, and J. Berri, "A Service-Based Mobile Tourist Advisor," *International Journal of Computer Information Systems and Industrial Management Applications,* vol. 1, pp. 177-187, 2009.

[8] V. Basili, G. Caldeira, and H. D. Rombach, "The Goal Question Metric Approach," *Encyclopedia of Software Engineering,* 1994.

[9] D. Zhang and B. Adipat, "Challenges, Methodologies, and Issues in the Usability Testing of Mobile Applications," *International Journal of Human-Computer Interaction,* vol. 18, pp. 293 - 308, 2005.

[10] B. Ballard, *User Interface Design Guidelines for J2me Midp 2.0*. Lawrence, KS, USA: Little Springs Design, Inc., 2005.

[11] J. Kjeldskov and J. Stage, "New techniques for usability evaluation of mobile systems," *International Journal of Human-Computer Studies,* vol. 60, pp. 599-620, 2004.

[12] S. Kristoffersen and F. Ljungberg, "Mobile informatics: innovation of IT use in mobile settings: IRIS'21 workshop report," *SIGCHI Bull.,* vol. 31, pp. 29-34, 1999.

[13] S. Björk, J. Redström, P. Ljungstrand, and L. E. Holmquist, "POWERVIEW - Using information links and information views to navigate and visualize information on small displays," in *Handheld and Ubiquitous Computing 2000 (HUC2k)* Bristol UK: Springer Verlag, 2000, pp. 46-62.

[14] S. Brewster, F. Chohan, and L. Brown, "Tactile feedback for mobile interactions," in *Proceedings of the SIGCHI conference on Human factors in computing systems* San Jose, California, USA: ACM, 2007.

[15] I. Minna, T. Marika, hti, V. Soili, in, m, and K. Kari, "Experimental evaluation of five methods for collecting emotions in field settings with mobile applications," *Int. J. Hum.-Comput. Stud.,* vol. 65, pp. 404-418, 2007.

[16] L. Barnard, J. S. Yi, J. A. Jacko, and A. Sears, "An empirical comparison of use-in-motion evaluation scenarios for mobile computing devices," *International Journal of Human-Computer Studies,* vol. 62, pp. 487-520, 2005.

[17] I. S. MacKenzie and R. W. Soukoreff, "Phrase sets for evaluating text entry techniques," in *CHI '03 extended abstracts on Human factors in computing systems* Ft. Lauderdale, Florida, USA: ACM, 2003.

[18] H. Dong-Han, H. Jeongyun, F. Peter, W. William, P. Sanghyun, S. Chiwon, and B. Mike, "Conceptual framework and models for identifying and organizing usability impact factors of mobile phones," in *Proceedings of the 20th conference of the computer-human interaction special interest group (CHISIG) of Australia on Computer-human interaction: design: activities, artefacts and environments* Sydney, Australia: ACM, 2006.

[19] P. Cairns and A. L. Cox, *Research methods for human-computer interaction*. Cambridge: Cambridge University Press, 2008.

[20] D. Scapin and E. Law, "Report, Review and Refine Usability Evaluation Methods (R3-UEM)," Athens 2007.

[21] A. Kaikkonen, A. Kekäläinen, M. Cankar, T. Kallio, and A. Kankainen., "Usability Testing of Mobile Applications: A Comparison between Laboratory and Field Testing," *Jurnal of Usability study,* vol. 1, pp. 4 -16, 2005.

[22] S. Ahmed, D. Mohammad, B. K. Rex, and K. P. Harkirat, "Usability measurement and metrics: A consolidated model," *Software Quality Control,* vol. 14, pp. 159-178, 2006.

[23] A. Danesh, K. Inkpen, F. Lau, K. Shu, and K. Booth, "GeneyTM: designing a collaborative activity for the palmTM handheld computer.," in *Proceedings of the SIGCHI conference on Human factors in computing systems* Seattle, Washington, United States ACM, 2001, pp. 388-395

[24] F. Erik, kj, H. Morten, and H. Kasper, "Measuring usability: are effectiveness, efficiency, and satisfaction really correlated?," in *Proceedings of the SIGCHI conference on Human factors in computing systems* The Hague, The Netherlands: ACM, 2000.

[25] J. Nielsen, *Usability Engineering*. London: Morgan Kaufmann, 1994.

[26] M. Ziefle, "The influence of user expertise and phone complexity on performance, ease of use and learnability of different mobile phones," *Behaviour and Information Technology,* vol. 21, pp. 303-311, 2002.

[27] R. Gafni, "Framework for Quality Metrics in Mobile-Wireless Information Systems," *Interdisciplinary Journal of Information, Knowledge, and Management,* vol. 3, pp. 23 - 38, 2008.

[28] L. Terrenghi, M. Kronen, and C. Valle, "Usability Requirements for Mobile Service Scenarios," in *Human Computer Interaction 2005 (IHCI'05),* Las Vegas, USA, 2005, pp. 1-10.

[29] G. Burnett, "USABLE VEHICLE NAVIGATION SYSTEMS: ARE WE THERE YET?," in *Vehicle Electronic Systems 2000 - European conference and*

exhibition Leatherhead, UK: ERA Technology Ltd, 2000.
[30] A. Blandford, P. Curzon, J. Hyde, and G. Papatzanis, "EMU in the Car: Evaluating Multimodal Usability of a Satellite Navigation System," in *Interactive Systems. Design, Specification, and Verification*, 2008, pp. 1-14.
[31] R. Solingen and E. Berghout, *The Goal/Question/Metric Method: A Practical Guide for Quality Improvement of Software Development*. Maidenhead, England: McGraw-Hill Publishing Company, 1999.
[32] S. Henry and D. Kafura, "Software Structure Metrics Based on Information Flow," *IEEE Trans. Softw. Eng.*, vol. 7, pp. 510-518, 1981.
[33] L. Buglione and A. Abran, "Balanced Scorecards and GQM: what are the differences?," in *FESMA/AEMES Software Measurement Conference*, Madrid, Spain, 2000.
[34] K. Heiko, "Goal, question, metric," in *Dependability metrics*, E. Irene, C. F. Felix, and R. Ralf, Eds.: Springer-Verlag, 2008, pp. 39-42.
[35] P. Berander and P. Jonsson, "A goal question metric based approach for efficient measurement framework definition," in *Proceedings of the 2006 ACM/IEEE international symposium on Empirical software engineering* Rio de Janeiro, Brazil: ACM, 2006.
[36] J. J. B. Vicente and F. Klett, "A Generic Evaluation Framework for Knowledge-Based Infrastructures: Design and Applications," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 3, pp. 290 -297, 2011.
[37] A. Hussain and M. Kutar, "Usability Metric Framework for Mobile Phone Application," in *The 10th Annual Conference on the Convergence of Telecommunications, Networking & Broadcasting*, Liverpool, UK, 2009, pp. 207 - 211.
[38] B. Kitchenham, "Procedures for undertaking systematic reviews.," Technical Report TR/SE-0401, Department of Computer Science, Keele University and National ICT, Australia Ltd, 2004. Joint Technical Report. 2004.
[39] L. Baillie, "Future Telecommunication: Exploring actual use," in *International Conference on Human-Computer Interaction, (INTERACT ´03)* 2003.
[40] A. Pirhonen, S. Brewster, and C. Holguin, "Gestural and audio metaphors as a means of control for mobile devices," in *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves* Minneapolis, Minnesota, USA: ACM, 2002.
[41] J. S. Dumas, & Redish, J. C., *A Practical Guide to Usability Testing.* Portland: Intellect Book, 1999.
[42] J. Nielsen and T. K. Landauer, "A mathematical model of the finding of usability problems," in *Proceedings of the INTERCHI '93 conference on Human factors in computing systems* Amsterdam, The Netherlands: IOS Press, 1993.
[43] J. P. Chin, V. A. Diehl, and K. L. Norman, "Development of an instrument measuring user satisfaction of the human-computer interface," in *Proceedings of the SIGCHI conference on Human factors in computing systems* Washington, D.C., United States: ACM, 1988.

Author Biographies

Azham H. received Master degree in 2000 and now he is pursuing his PhD degree. He is currently working as lecturer in Faculty of Information Technology, Northern University of Malaysia. His current research interest includes the area of software metric and mobile technology.

M. Kutar received her PhD degree in Computer Science in 2001. She is currently a Senior Lecturer in Salford Business School, University of Salford, UK. Her research interest includes the effectiveness of representations used in system development, temporal aspects of interactive systems, psychology of programming, privacy, and law / ethics in IT.